العنوان:	التنبؤ بالمبيعات بإستخدام شجرة القرار: بالتطبيق على شركة سويتش للتجارة والهندسة قسم المولدات
المؤلف الرئيسي:	عبدالله، سمية عبدالله محمد الأمين
مؤلفين آخرين:	أحمد، شاذلى صديق محمد(مشرف)
التاريخ الميلادي:	2017
موقع:	الخرطوم
الصفحات:	1 - 82
رقم MD:	916054
نوع المحتوى:	رسائل جامعية
اللغة:	Arabic
الدرجة العلمية:	رسالة ماجستير
الجامعة:	جامعة النيلين
الكلية:	كلية علوم الحاسوب وتقانة المعلومات
الدولة:	السودان
قواعد المعلومات:	Dissertations
مواضيع:	التنبؤ بالمبيعات، شركة سويتش للتجارة والهندسة، قسم المولدات، تقنية أشجار القرار، التنقيب عن البيانات، قواعد البيانات
رابط:	http://search.mandumah.com/Record/916054

© 2024 المنظومة. جميع الحقوق محفوظة.
هذه المادة متاحة بناء على الإتفاق الموقع مع أصحاب حقوق النشر، علما أن جميع حقوق النشر محفوظة. يمكنك تحميل أو طباعة هذه المادة للاستخدام الشخصي فقط، ويمنع النسخ أو التحويل أو النشر عبر أي وسيلة (مثل مواقع الانترنت أو البريد الالكتروني) دون تصريح خطي من أصحاب حقوق النشر أو المنظومة.

للإستشهاد بهذا البحث قم بنسخ البيانات التالية حسب إسلوب الإستشهاد المطلوب:

إسلوب APA
عبدالله، سمية عبدالله محمد الأمين، و أحمد، شاذلى صديق محمد. (2017).التنبؤ بالمبيعات بإستخدام شجرة القرار: بالتطبيق على شركة سويتش للتجارة والهندسة قسم المولدات(رسالة ماجستير غير منشورة). جامعة النيلين، الخرطوم. مسترجع من http://search.mandumah.com/Record/916054

إسلوب MLA
عبدالله، سمية عبدالله محمد الأمين، و شاذلى صديق محمد أحمد. "التنبؤ بالمبيعات بإستخدام شجرة القرار: بالتطبيق على شركة سويتش للتجارة والهندسة قسم المولدات" رسالة ماجستير. جامعة النيلين، الخرطوم، 2017. مسترجع من http://search.mandumah.com/Record/916054

© 2024 المنظومة. جميع الحقوق محفوظة.
هذه المادة متاحة بناء على الإتفاق الموقع مع أصحاب حقوق النشر، علما أن جميع حقوق النشر محفوظة. يمكنك تحميل أو طباعة هذه المادة للاستخدام الشخصي فقط، ويمنع النسخ أو التحويل أو النشر عبر أي وسيلة (مثل مواقع الانترنت أو البريد الالكتروني) دون تصريح خطي من أصحاب حقوق النشر أو المنظومة.

جامعة النيلين

كلية الدراسات العليا

كلية علوم الحاسوب وتقانة المعلومات

بحـــث تكميلي لنيل درجـــة المــاجســـتيرفي تقانة المعلومات

بعنـــوان:

التنبؤ بالمبيعات باستخدام شجرة القرار

(بالتطبيق على شركة سويتش للتجارة والهندسة قسم المولدات)

إعداد الطالبة:

سمية عبد الله محمد الامين عبد الله

إشراف الدكتور

شاذلي صديق محمد احمد

2016 - 2017 م

بسم الله الرحمن الرحيم

الآيات

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

الرَّحْمَنُ (1) عَلَّمَ الْقُرْآنَ (2) خَلَقَ الْإِنْسَانَ (3) عَلَّمَهُ الْبَيَانَ (4) الشَّمْسُ وَالْقَمَرُ بِحُسْبَانٍ (5) وَالنَّجْمُ وَالشَّجَرُ يَسْجُدَانِ (6) وَالسَّمَاءَ رَفَعَهَا وَوَضَعَ الْمِيزَانَ (7) أَلَّا تَطْغَوْا فِي الْمِيزَانِ (8)

صدق الله العظيم

"سورة الرحمن الأية 1-8"

الأهداء

الى

امي التي عانت وجاهدت لاكون انا ...

والدي الذي كابد عناء الحياة لنتعلم ونسمو...

الى اخواتي العزيزات سامية,سلوى وسارة تؤائم روحي واشقائي الاعزاء

الي زملائي الذين حقا استفدت وتعلمت من رفقتهم لكم اساتذتي الاجلاء

ولكل من علمني حرفاً ....

شكر وعرفان

لابد لي وانا اخطو خطواتي الأخيرة في نيل درجة الماجستير من وقفة اعود فيها إلى الاوقات التي قضيتها في رحاب جامعة النيلين مع أساتذتي الكرام الذين قدموا لي الكثير واجتهدوا في توصيل المعلومة وكل ما هو جديد...

وقبل أن أمضي أقدم أسمى آيات الشكر والامتنان والتقدير والمحبة إلى الذين حملوا أقدس رسالة في الحياة إلى الذين مهدوا لي طريق العلم والمعرفة.

إلى جميع أساتذتي الأفاضل وأخص بالتقدير والشكر الدكتور **شاذلي صديق محمد احمد** الذي بذل صبرا ليرى بحثي هذا النور والدكتور**طارق عبد الكريم** الذي علمنا ماهو التنقيب في البيانات وكل الدكاترة الذين افادونا علما.

وكذلك اشكر إدارة شركة سويتش للتجارة والهندسة ومديرها العام **الاستاذ احمد الرشيد سعد** والزملاء الاعزاء الذين لم يبخلوا علي بالبيانات و كل من ساعدني على إتمام هذا البحث وقدم لي العون ومد لي يد المساعدة وزودني بالمعلومات اللازمة لإتمام هذا البحث

نسأل الله أن يوافيهم بأحسن ما عملوا ويجزيهم خير الجزاء.

المستخلص

مع وجود كميات كبيرة من البيانات المخزنة في قواعد البيانات و مخازن البيانات ازدادت الحاجة إلى تطوير أدوات تمتاز بالقوة لتحليل البيانات و استخراج المعلومات و المعارف منها، من هنا ظهر ما يسمى بالتنقيب في البيانات كتقنية تهدف إلى استخراج المعرفة من كميات هائلة من البيانات، و هي تقنية حديثة فرضت نفسها بقوة في عصر المعلوماتية(عصر الانترنيت و الاقتصاد الرقمي) الذي يتميز بالسيل العظيم و الانتشار الواسع النطاق للبيانات حتى أضحى من المستحيل على المحللين استخلاص معلومات ذات معنى باللجوء فقط إلى المداخل التقليدية للتحليل التمهيدي للبيانات

عناصر التسويق الاساسية هي المنتج,التوزيع او المكان, الاعلان والسعر وبما ان شركة سويتش تدفع مبالغ طائلة في الاعلانات ويكون العائد دون الطموح اي انه غير متناسبا مع ما قدم من جهد وذلك لان الاعلان غير موجه وبالتالي تهدر الاموال دون الوصول الى القصد كما ان شركة سويتش من الشركات الأوائل علي نطاق السودان في مجال المولدات ولكن بد ظهور شركات منافسة كان لابد من معرفة متطلبات السوق وهذا لا يتاتى الا بالدراسة للمعلومات طوال السنين السابقة حتى نعرف احجام المولدات الاكثر طلبا (اي المنتج) ويمكن بذلك معرفة سلوك العميل كما يمكن من خلال هذه الدراسة الاستمرار لمعرفة الحوجة الحقيقية للعميل وتنويره بالمناسب من حيث سعة المنتج (المولد) وفتح سوق جديد للمناطق الجغرافية الغير مغطاه .

استخدمنا هنا التصنيف وهو احدى تقنيات التعدين في تفسير أو التنبؤ بخاصية ما من خلال خصائص أخرى وذلك باستخدام بيانات عملاء شركة سوتش المخزنة والتي تم جمعها مسبقا من معلومات العميل وتصنيفها تصنيفا وصفي تحليلي والاداة المستخدمة هي weka على حسب خاصية محددة هي سعة المنتج (المولد) كانت الدقة 98.82 للسعات large ونسبة الخطأ 1.18

يوصى في المستقبل التركيز في التسويق على المنتج(المولد) بالسعة portable الذي اتضح ان شرائه ضعيف .

كما يوصى بمراعاه ادخال البريد الالكتروني كعنصر اساسي داعم للتسويق لانه القاسم المشترك لمعظم وسائل التواصل الاجتماعي وكذلك التلفون

Abstract

With the existence of large amounts of data stored in databases and data warehouses, the need to develop tools that have the power to analyze data and extract information and knowledge has increased. Hence, the so-called datamining has emerged as a technique aimed at extracting knowledge from vast amounts of data. A modern technology that has strongly established itself in the era of information technology (the Internet age and the digital economy), characterized by the great torrent and widespread data, it has become impossible for analysts to extract meaningful information by resorting only to traditional approaches to data analysis

The basic marketing elements are the product, the distribution, the place, the advertisement and the price. Since SWITCH pays huge sums in advertising and the return is without ambition, it is not proportional to the effort made because the advertising is not directed and thus waste the money without reaching the intention. Of the first companies in the field of Sudan in the field of generators, but the emergence of competing companies had to know the requirements of the market and this is not only the study of information over the previous years so as to know the sizes of generators most requested and can thus know the behavior of the client can also through this study to continue to know the need To the client and enlighten him appropriately in terms of the capacity of the product (generator) and open a new market for geographical areas not covered.

Here we used the classification, which is one of the techniques of mining in the interpretation or prediction of a property through other properties, using the data of the clients stored company and stored previously collected from the customer information and classification descriptive analytical and the tool used by weka according to a specific property is the capacity of the product (generator) 98.82 for large and error ratio of 1.18

It is recommended to focus on marketing on a portable capacitive product that turned out to be weak.

It is also recommended that e-mail be included as a key element supporting marketing because it is the common denominator of most social media as well as telephone

فهرست المحتويات

رقم الصفحة	الموضوع
أ	الأيات من 1-8 سورة الرحمن
ب	الإهداء
ج	شكر وعرفان
د	المستخلص
هـ	Abstract
و	فهرست المحتويات
ط	فهرست الجداول
ي	فهرست الاشكال
	الفصل الاول(الإطار العـــــــام)
1	1-1المقدمة
2	1-2مشكلة البحث
2	1-3أهداف البحث
2	1-4اهمية البحث
3	1-5منهجية البحث
3	1-6حدود البحث
3	1-7هيكلية البحث
	الفصل الثاني(الإطار النظري)
5	2-1الدراسات السابقة
10	2-1-4مقارنة بين الدراسات السابقة والدراسة الحالية
12	2-2-1مخازن البيانات(Data Warehousing)
14	2-2-2 اهداف التنقيب
14	2-2-3وسائل التنقيب في البيانات

2-2-4طرق ونسق التنقيب في البيانات	16
2-2-5الفرق بين عملية الاستعلام العاديه على قواعد البيانات العاديه والاستعلام في البيانات المنقبه	16
2-2-6.تهيئة البيانات (Data preparation)	17
2-2-7-1تنظيف البيانات(data cleaning)	17
2-2-7-2 البيانات المفقودة	18
2-2-7-3 اشتقاق البيانات	18
2-2-7-4 دمج البيانات	18
2-2-8تطبيقات التنقيب في البيانات	19
2-2-9أمثلة عن أهمية تطبيق تقنيات التنقيب في البيانات	19
2-2-10كيفية دراسة البيانات	20
2-2-11عملية التنقيب	22
2-2-12 مراحل اكتشاف المعرفة	26
2-3-1أهم طرق تنقيب البيانات	29
2-3-2التصنيف Classification	29
2-3-3خطوات التصنيف	29
2-3-4طريقة تعلم البيانات	29
2-3-5تقسيم البيانات	30
2-3-6مقاييس اختيار الصفة	30
2-3-7 تقنيات التصنيف	33
2-3-7-1تقنية أشجار القرار Decision Trees	33
2-3-7-2 خصائص تقنية شجرة القرار	34
2-3-7-3مميزات شجرة القرارات	34
2-3-7-4عيوب **شجرة القرارات**	34
2-3-7-5 التصور Visualization	34

2-4 سوق البيانات (DATA MARTS)	36
2-4-1ماهو سوق البيانات (data mart)	36
2-4-2الاختلافات الأساسية بين مستودع البيانات وسوق البيانات	36
2-4-3الإدارة الذكية والتسويق	37
2-4-4الادارة الذكية	37
2-4-5التسويق الذكي واستهداف العملاء	39
2-5برنامج Weka	42
2-6 نبذة عن شركة سويتش للتجارة والهندسة	49
الفصل الثالث (تحليل الخوارزميات)	
3-1التحليل	50
3-2تعريف المشكلة (تحليل المشكلة)	50
3-3مرحلة جمع البياناتData Collection	50
3-4 وصف البيانات Dataset Description	51
3-5البرنامج المستخدم	51
3-6 تجهيز البيانات Dataset processing	53
3-7 تنقية البيانات dataset Filter	53
3-8 شرح خطوات الخوارزمية شجرة القرار (J48)	53
3-9 استخدام خوارزمية شجرة القرار في التحليل	54
الفصل الرابع (التصميم والتطبيق)	
4-1التصميم	56
4-2 التطبيق	61
الفصل الخامس(النتائج والتوصيات)	
5-1 النتائج	71
5-2النتائج النهائية	77
5-3التوصيات	78

4-5الخاتمة	79
5-5 المراجع والمصادر	80

فهرست الجداول والاشكال

الجداول

رقم الجدول	عنوان الجدول	رقم الصفحة
جدول رقم (1)	مقارنة بين الدراسات السابقة والدراسة الحالية	10
جدول رقم (2)	امثلة لبعض وسائل التنقيب	15
جدول رقم(3)	تنظيف البيانات (Data cleaning)	18
جدول رقم (4)	مقارنة بين مخازن البيانات (data warehouse) وسوق البيانات (data mart)	36
جدول رقم (5)	خواص ومواصفات كل متغير	51
جدول رقم (6)	مجموعة البيانات لمبيعات شركة سويتش 2015-2017 في اكسل	52
جدول رقم(7)	البيانات قبل تحويلها الى ملف بالامتداد csv.*	57
جدول رقم (8)	البيانات بعد تحويلها الى ملف بالامتداد csv.*	58

الاشكال

رقم الشكل	عنوان الشكل	رقم الصفحة
شكل رقم (1)	تنقيب البيانات وذكاء الاعمال Data mining and Business intelligence	12
شكل رقم(2)	تهيئة البيانات**data preparation**	17
شكل رقم (3)	تمثيل القيمة المفقودة	18
شكل رقم (4)	توضيح عملية تنقيب النصوص باستخدام التجميع	24
شكل رقم (5)	رسم يوضح مراحل اكتشاف المعرفة	26
شكل رقم (6)	شكل يوضح اساليب تعدين البيانات	28
شكل رقم (7)	شكل يوضح شجرة القرار	33
شكل رقم (8)	شكل يوضح التصنيف	35
شكل رقم (9)	WEKA واجهة الاداة مفتوحة المصدر	43
شكل رقم (10)	عملية تصنيف بواسطة برنامج weka	45
شكل رقم (11)	نسبة الحالات الصحيحة في المائة و نسبة الحالات الخطأ في المائة	46
شكل رقم (12)	مصفوفة التضارب	46
شكل رقم (13)	مصفوفة التضارب (الفئات المتنبئة والفئات الفعلية)	47
شكل رقم (14)	نسبة الصحة والخطأ والدقة	48
شكل رقم (15)	المخطط الانسيابي لخوارزميةJ48(decision Tree)	54
شكل رقم (16)	مرحلة التصميم الشاشة الرئيسية لبرنامج weka	56
شكل رقم (17)	شاشة المعالجة الخاصة ببرنامج الويكا	57
شكل رقم (18)	فتح ملف switchgroup.csv في برنامج الويكا	60
شكل رقم (19)	مرحلة إضافة مجموعة البيانات الى برنامج الweka	61

شكل رقم (20)	مرحلة المعالجة الاولية للبيانات	62
شكل رقم (21)	نسب المبيعات	63
شكل رقم (22)	تنفيذ خوارزمية التصنيف على البيانات الخاصة بمبيعات سوتش 2015_2017(تفاصيل البيانات المدخلة)	64
شكل رقم (23)	تنفيذ خوارزمية التصنيف على البيانات الخاصة بمبيعات سوتش 2015_2017(معلومات شجرة القرار الناتجة)	65
شكل رقم (24)	تنفيذ خوارزمية التصنيف على البيانات الخاصة بمبيعات سوتش 2015_2017(شكل شجرة القرار الناتجة)	66
شكل رقم (25)	تنفيذ خوارزمية التصنيف على البيانات الخاصة بمبيعات سوتش 2015_2017(ملخص النتائج)	67
شكل رقم (26)	تنفيذ خوارزمية التصنيف على البيانات الخاصة بمبيعات سوتش 2015_2017(نسبة الصحة والخطأ والدقة)	68
شكل رقم (27)	تنفيذ خوارزمية التصنيف على البيانات الخاصة بمبيعات سوتش 2015_2017(مصفوفة التضارب)	69
شكل رقم (28)	التصور Visualize	70
شكل رقم (29)	مناقشة النتائج (معلومات التنفيذ)	71
شكل رقم (30)	نموذج المصنف	72
شكل رقم (31)	تقييم المصنف على بيانات الاختبار	72
شكل رقم (32)	مصفوفة التضارب	73
شكل رقم (33)	مصفوفة التضارب الحقيقي والمتنبا به	74
شكل رقم (34)	النتائج النهائية	75

الفصل الاول

(الإطار العـام)

1-1المقدمة

مع وجود كميات كبيرة من البيانات المخزنة في قواعد البيانات و مخازن البيانات ازدادت الحاجة إلى تطوير أدوات تمتاز بالقوة لتحليل البيانات و استخراج المعلومات و المعارف منها، من هنا ظهر ما يسمى بالتنقيب في البيانات كتقنية تهدف إلى استخراج المعرفة من كميات هائلة من البيانات، و هي تقنية حديثة فرضت نفسها بقوة في عصر المعلوماتية(عصر الانترنيت و الاقتصاد الرقمي) الذي يتميز بالسيل العظيم و الانتشار الواسع النطاق للبيانات حتى أضحى من المستحيل على المحللين استخلاص معلومات ذات معنى باللجوء فقط إلى المداخل التقليدية للتحليل التمهيدي للبيانات ، و استخدام التنقيب في البيانات يوفر للشركات و المنظمات في جميع المجالات القدرة على الاستكشاف و التركيز على أهم المعلومات في قواعد البيانات، كما تركز تقنيات التنقيب على بناء التنبؤات المستقبلية و استكشاف السلوك و الاتجاهات مما يسمح باتخاذ القرارات الصحيحة و اتخاذها في الوقت المناسب، و التي تعتبر بدورها مرحلة من مراحل عملية أكثر تعقيدا هي استكشاف المعرفة في قواعد البيانات، و المرتبطة إلى حد بعيد بعملية تطوير أخرى مهمة جدا هي مستودعات البيانات. حيث أن الكثير من الشركات و المنظمات الرائدة اليوم تستخدم عملية استكشاف المعرفة في قواعد البيانات بشكل منهجي و منظم بوصفها تشكل جوهر العمل الذي يعتمد عليه في تفعيل النشاط و تحقيق الميزة التنافسية [1].

يستخدم التصنيف وهو احدى تقنيات التعدين في تفسير أو التنبؤ بخاصية فرد ما من خلال خصائص أخرى هذه الخاصية هي عموما كيفية وليست كمية و يمكن انجاز التصنيف بالاعتماد على الأساليب الإحصائية القديمة مثل الانحدار و التحليل التمييزي، أو بالاعتماد على أساليب حديثة نسبيا مثل قوى الارتباط و الاستنتاج المستند إلى الحالة و الشبكات العصبية و كأمثلة عن طرق التصنيف المستعملة كجزء من تطبيقات استكشاف المعرفة التي تتضمن تصنيف اتجاهات الأسواق المالية، و التحديد الآلي للأشياء المهمة في صورة كبيرة من قواعد البيانات.[2]

سوق البيانات data marts وهو شكل بسيط من مستودع البيانات(data ware house) ولكنه يركز على موضوع واحد (أو مجال وظيفي) مثل المبيعات أو المالية أو التسويق وغالبا ما يتم بناء data marts والتحكم بها من قبل إدارة واحدة داخل المنظمة .وبالنظر إلى تركيزها على موضوع

واحد، فإن البيانات عادة ما تستمد من مصادر قليلة فقط ويمكن أن تكون المصادر أنظمة تشغيل داخلية أو مستودع بيانات مركزي أو بيانات خارجية. [3]

1-2مشكلة البحث :

وجود كم هائل من البيانات لدى قسم التسويق خاص بمبيعات السنين السابقة وعدم وجود تقنية للاستفادة من تلك البيانات لمعرفة سلوك الزبائن الشرائية .

1. ايجاد تقنية للتنبؤ بمتطلبات السوق في المستقبل حسب سعات المنتج

1-3أهداف البحث:

1. توجية المؤسسة للمنتج المستهلك (المرغوب)
2. معرفة المنتجات ذات الشراء الضعيف حتى يتم التسويق لها بصورة صحيحة
3. ترشيد موارد المؤسسات باستخدام تقنية التصنيف.
4. جعل التسويق فعال بإتباعه طرق فعالة
5. الاستفادة من بيانات المبيعات المخزنة سابقا

1-4اهمية البحث:

هذا البحث مهم في حفظ موارد الجهة المقدمة للتسويق _هنا شركة سويتش للتجارة والهندسة _ من جهد ومال ووقت والاستفادة من البيانات المخزنة مسبقا ، ومن ناحية اخرى فهو مهم لتحديد المنتج الاكثرا مبيعا في المستقبل والاقل مبيعا حسب تصنيف البيانات الموجودة حاليا بسوق البيانات كما يمكن معرفة الاسباب وتركيز عملية الاعلان علي ضعيف الشراء.

1-5منهجية البحث :

الوصفي التحليلي والاداة المسنخدمة هي weka واستخدام J48 لتنفيذ خوارزمية شجرة القرار C4.5

1-6حدود البحث :

الحدود المكانية :شركة سويتش للتجارة والهندسة

الحدود الزمانية:2015-2017

1-7هيكلية البحث:

ينقسم هذا البحث إلي خمسه فصول:

الفصل الاول (الاطار العام) ويحتوى علي المقدمة-المشكلة-الأهداف-اهمية البحث - المنهجية والأدوات - حدود البحث-هيكل البحث

الفصل الثاني(الاطار النظري) ويشمل الدراسات السابقة - تعدين البيانات (مراحل اكتشاف المعرفة) - DATA MARTS التصنيف (استخدامه في التسويق) - خلفية عن البرنامج المستحدم WEKA

الفصل الثالث(تحليل الخوارزميات) ويحتوي علي نبذة عن شركة سويتش - مرحلة جمع البيانات(المصادر) _ توضيح وسيلة جمع المعلومات ووصف البيانات نفسها - تعريف المشكلة (تحليل المشكلة) - شرح خطوات الخوارزمية بصورة تقنية - رسم المخطط الانسيابي للخوارزمية

الباب الرابع(النتائج والتوصيات) ويحتوي علي التصميم والتطبيق - التطبيق

والفصل الخامس النتائج والتوصيات يشمل مناقشة النتائج - التوصيات - المراجع والمصادر

الفصل الثاني

(الاطار النظري)

2-1الدراسات السابقة

2-1-1الدراسة الاولى بواسطة الطالبة / داليا عبد الحسين احمد دولة العراق بحث لنيل درجة الماجستير من م.م /الجامعة التكنولوجية بعنوان **دور تنقيب البيانات في زيادة أداءالمنظمة**(دراسة تحليليه في المصرف الصناعي ببغداد) إشراف أ .م .د .زكريا مطلك الدوري جامعة بغداد /كلية الادارة والاقتصاد المجلد/13ع/ 48 لسنة2007 .

تتلخص مشكلة البحث في ثلاث نقاط أساسية وجود عدد كبير من المقترضين في المصرف الصناعي ,وجود كم هائل من البيانات عنهم وعدم وجود تقنية تمكن المصرف من اتخاذ قرار سريع.

كما يمكن صياغة بعض التساؤلات مثل:ما مدى اهتمام المصرف في التوثيق الكامل للبيانات؟,كيف تستخدم الأساليب العلمية والتكنولوجية في التعامل مع هذه البيانات؟,هل تبنى اتخاذ القرارات على ما يستخلص من البيانات ام على الحدس؟,هل يجري تقييم هذه البيانات وإكمالها وتحديثها باستمرار؟,هل ان استخدام تنقيب البيانات يزيد من اداء المصرف؟

أهمية البحث تتمثل في تناول البحث لمتغيرين مهمين تنقيب البيانات DM والاداء يمكن ان يجعل البحث مساهمة متواضعة في هذا المجال خاصة بعد ما لوحظ من نقص كبير جدا في المصادر العربية الخاصة بالمتغير الاول والذي يعد من الموضوعات الحرجة والحديثة جدا "حيث يلاحظ ان اقدم مصدر في هذا الموضوع قد يعود الى عام 1996 كحد أدنى،يساعد البحث المصرف على تجنب مخاطر حقيقية ناتجة عن عدم التسديد والتي تشكل عبئا"على المصرف لكونها قروضا "هالكة،تقديم نماذج تسهل عملية تنقيب البيانات في المصرف حيث تتنبأ بالمقترضين وتصنفهم,حسب مخاطر التسديد لكي يتمكن المصرف من التركيز على المقترضين الموثوق بهم فقط,سعيا منه لزيادة الاداء,يقدم البحث للمصرف قاعدة بيانات يفتقر اليها تخص القروض والمقترضين كما يقدم نظام,متكامل لادارة قواعد البيانات يتمثل بالبرنامج الحاسوبي المستخدم.

يهدف هذا البحث إلي دراسة مشكلة تراكم البيانات ومحاولة إيجاد الحل الأمثل لها،محاولة تحسين مستوى استخدام المعلومات من قبل الادارات من اجل زيادة الاداء,بناء نماذج تنقيب البيانات التي يمكن ان تستخدم في التنبؤ بالمقترضين الموثوق بهم واستبعاد غير الموثوقين منهم لتقليل تكاليفهم و محاولة إرساء أسس سليمة تكون البداية لانطلاق بحوث لاحقة تحاول التطبيق في مجالات اخرى جديدة.

المنهجية المستخدمة: يعتمد البحث على تطبيق برنامج حاسوبي جاهز ومن ثم الوصف والتحليل لنتائج التطبيق من نماذج لاستخلاص المؤشرات الأساسية التي تدعم عملية اتخاذ القرار والرجوع مرة اخرى للتحليل للوقوف على امتدادات نتائج التطبيق.

النتائج التي توصل إليها الباحث هي:

اولا-"الاستنتاجات المتعلقة بالجانب النظري

1. توفر عملية تنقيب البيانات ميزة التعامل مع كميات هائلة من البيانات وكلما كانت كمية البيانات كبيرة كانت اكثر قدرة ودقة في الوصول الى نماذج صحيحة مفيدة.
2. يختلف تنقيب البيانات عن أنظمة دعم القرار كونه قد يكشف عن علاقات ابعد من ان يفكر بها المستخدم.
3. تعتمد نتائج تنقيب البيانات بصورة اساسية على مدى دقة وصحة وشمولية البيانات المستخدمة.
4. تستخدم في تنقيب البيانات تقنيات وادوات عديدة ومتنوعة لكنها تصب في اتجاهين أساسيين هما الاستكشاف والتنبؤ.

لا يمكن تحديد خوارزمية معينة باعتبارها الافضل لبناء نموذج تنقيب البيانات بل يفضل ان تستخدم خوارزميات متعددة ويتم بناء نماذج متعددة الى ان يتم الوصول الى النموذج الافضل.

5. في حالة وجود قاعدة بيانات كبيرة جدا، يكون خيار بناء نماذج عديدة مبنية على عينات افضل من خيار بناء نماذج قليلة مبنية على كل البيانات لكي يكون النموذج اكثر دقة ومتانة.
6. تتطلب عملية قياس الاداء اختيار مؤشرات واضحة سهلة الاستخدام وقليلة الوقت والجهد

ثانيا-"الاستنتاجات المتعلقة بالجانب التطبيقي

1. كلما كان عدد المتغيرات او الاعمدة في قواعد البيانات اكثر كان نموذج التنقيب الناتج اكثر
2. واقعية في التعامل مع مشاكل العمل.
3. سهولة تنقيب البيانات من مصادر بيانات علائقية .SQL Server 2 يوفر برنامج2000
4. ومتعددة الابعاد ومن خلال خوارزميتين هما شجرة القرار والتجميع.
5. يمكن اجراء عملية التنقيب على بيانات تضم متغيرات رقمية وغير رقمية وتظهر في
6. النموذج بالضبط كما صيغتها في قاعدة البيانات مثل مستوى القرض الذي يحمل تقديرات مثل واطىء ومتوسط وعالي.

7. تكون النماذج الناتجة عن عملية التنقيب وخاصة من البرنامج المستخدم سهلة الفهم حيث يستطيع

المستخدم مثل المحلل المالي ان يفسر النموذج ويقارنه مع افكاره الخاصة ,ان سبب ذلك يعود الى ان البرنامج يعمل على اختبار نماذج عديدة ويقدم في النهاية النموذج

8. وفر تطبيق البرنامج عامل السرعة حيث يمكن الوصول الى نموذج التنقيب في غضون دقيقة او اثنتين كحد أقصى مع السماح بتحديث هذه النماذج باستمرار ,ان هذا التحديث والسرعة يسهل على ادارة المصرف اتخاذ القرار الخاص بالتعامل مع طالبي القروض بالرفض او القبول.

9. نتج عن استخدام خوارزمية شجرة القرار نموذج تنبؤي حدد فيه مستوى القرض Level الصفة الأكثر أهمية في التنبؤ بمخاطر القروض ويكون مستوى القرض هنا اما واطئا" او غير واطئ واختارت الخوارزمية بعد ذلك الفرعBranch الصفة الأكثر أهمية الثانية وهو اما ان يكون فرع المصرف في بابل او لا يكون كذلك ويصاحب هذا النموذج وصف واضح ومبسط لكل اجزائه.

10. ينتج عن استخدام خوارزمية التجميع نموذج وصفي حدد عشرة تجمعات مرتبة تنازليا حسب عدد القيود من التجمع الاول وحتى التجمع العاشر، حيث يضم مثلا التجمع السابع اكبر عدد من الحالات السيئة للقروض تصل الى 28 حالة خطر سيئة مقابل 2 حالة خطرة جيدة فقط مع نسب احتمالية كل منها وتنحصر مخاطر القروض السيئة في فروع المصرف في اربيل والسليمانية والمستثمرة في الصناعات الانشائية والميكانيكية ومحطات تعبئة الوقود والنجارة وتكون مستويات القروض متوسط وعالية والمقترضين هم افراد من كلا الجنسين ذكور واناث.

11. ثبوت صحة الفرضية الاساسية حيث ان تطبيق النماذج الناتجة عن خوارزميتي شجرة القرار والتجميع ، يمكن الادارة من رفع نسبة ما مسدد الى ما هو مستحق من قروض ويؤدي ذلك الى تقليل الكلف وزيادة صافي الربح وهو ما يظهر جليا في معيار صافي الربح/ رأس المال حيث حقق المصرف اعلى المعدلات في سنة 2001 ووصل الى699% تقابلها نسبة تسديد عالية نوعا ما مقارنة ببقية السنوات ووصلت الى50.7%

2-1-2 الدراسة الثانية بعنوان بناء **شجرة القرار باستخدام خوارزمية C4.5 لدعم قرارات التسويق المباشر في المصارف لنيل درجة الماجستير بتاريخ 2016/6/13 من جامعة حلب بواسطة** الطالب محمد حسن صديق قسم الاحصاء كلية الاقتصاد جامعة حلب اشراف الدكتور مروان مصطفى ناعسة

مشكلة البحث: تتمثل مشكلة البحث الاساسية في عدم قدرة الاساليب التقليدية على تصنيف العملاء المستهدفين في الحملات التسويقية المباشرة هذا يعني عدم القدرة على التنبؤ باستجابة العملاء الى هذه الحملات وهذا يطرح التساؤلات التالية :

1. هل يمكن استخدام خوارزمية C4.5 في بناء شجرة قرارات تعمل على تصنيف العملاء حول استجابتهم للحملات التسويقية المباشرة ؟
2. ماهو مستوى الإعتماد على هذه الخوارزمية في التنبؤ باستجابة العملاء للحملة التسويقية ؟

اهمية البحث: تاتي اهمية البحث من خلال استخدام تقنية التصنيف كاحدى تقنيات التنقيب في البيانات في تصنيف عملاء المصرف وبالتالي استخلاص قواعد وانماط غير معروفة من كميات كبيرة من البيانات والتي لايمكن اكتشافها بالطرق التقليدية ، والتنبؤ باستجابة العملاء للحملات التسويقية المباشرة مما يزيد في الربحية والحصة السوقية للمصارف من جهة وتخفيض تكاليف الاتصال والوقت من جهة اخرى ,كم تنتج اهمية البحث بتسليط الضوء على اهم طرق التقييم التي يمكن بها الحكم على التقنية المستخدمة ,والاعتماد عليها في الحملات التسويقية المباشرة في المصارف بشكل فعلي وقابل للتطبيق

اهداف البحث: يهدف الى :

1. بناء نموزج شجرة القرار بإستخدام خوارزمية C4.5 ,تعمل على تصنيف العملاء في الحملات التسويقية المباشرة في المصارف
2. تقييم النموزج وفق عدة مقاييس , واختيار قدرته على الوصول للعملاء الذين سوف يستجيبون للحملات التسويقية قبل الاعتماد عليه بشكل فعلي .

منهجية البحث: يعتمد البحث على المنهج التحليلي, حيث يتم تحليل البيانات الخاصة بالعملاء باستخدام الاداة weka لبناء نموزج شجرة قرار للتنبؤ باستجابة العملاء للحملات التسويقية المباشرة في المصارف باستخدام خوارزمية C4.5 ,وسيتم عبر عملية التدريب على جزء من من البيانات تكون فيها حالات الاستجابة اقل بكثير من حالات عدم الاستجابة بغرض تدريب النموزج على البحث عن حالات الاستجابة الايجابية للحملة التسويقية المباشرة ,وتقييم نتائج عملية التدريب ثم بعد ذلك تت عملية اختبار النموذج على جزء آخر من البيانات ,حيث تكون فيها حالات الاستجابة وعدم الاسجابة متساوية بغرض معرفة قدرة النموذج على التفريق بين الحالتين وتقييم نتائج الاختبار للتاكد من دقة النموذج ومدى امكانية الاعتماد عليه في حملات مستقبليه.

النتائج التي توصل اليها الباحث:

1. كفاءة خوارزمية C4.5 على بناء شجرة قرار بغرض دعم القرارات التسويقية المباشرة في المصارف.
2. امكانية تدريب خوارزمية C4.5 على مجموعة من السجلات في ظروف اختلاف كميات الاصنافقي المتغير الهدف

3. القدرة على بناء نموزج كامل لبئية تنبؤية عالية ,للتنبؤ باستجابة العملاء للحملات التسويقية.

التوصيات :

1. نوصي المصارف بإستخدام خوارزمية C4.5 في بناء شجرة القرار لتصنيف العملاء حول الاستجابة للحملات التسويقية ,وذلك عبر تكوين قاعدة بيانات كاملة حول العملاء فالظروف المحيطة بشكل مسبق واستنتاج الانماط التي تحدد سلوكهم تجاه الحمله المستقبلية .
2. استخدام خوارزميات اخرى في بناء نموزج شجرة القرارات لدعم قرارات التسويق المباشر في المصارف وتقييم قدرة النموذج على التطبيق الفعلي.

3-1-2 الدراسة الثالثة ورقة علمية بعنوان استخدام تقنيات تنقيب البيانات لكشف التطفل في شبكات الحاسوب جمهورية السودان , جامعة النيلين، كلية علوم الحاسوب وتقانة المعلومات , الباحثان

أ.إياد محمد مهيوب غالب البريهي – باحث أول

أ.د السماني عبد المطلب أحمد – باحث ثاني

تهدف هذه الورقة لاستخدام تقنيات تنقيب البيانات data mining لكشف حالات الشذوذ في الشبكات تطبيقا على عينتين عشوائيتين من مجموعة بيانات NLS-KDD Data Set ، استخدمت الورقة تقنية التصنيف مصنف شجرة القرار decision tree التي تنفذ خوارزمية C4.5، نتيجة التجربة تعرض أن مصنف C4.5أظهر نتائج فعالة لكشف التطفل في الشبكات الحاسوبية وأظهرت النتائج أنه كلما كانت كمية البيانات كبيرة تكون نسبة الخطأ أقل ودقة التنبؤ عالية , حيث كانت دقة المصنف 63.9 %عندما كانت كمية بيانات الاختبار (946,848 byte) وعندما زادت كمية بيانات إلى (497,700 byte)ارتفعت دقة المصنف إلى (81.5339%) ، بينما كانت كمية بيانات التدريب ثابتة، هذه النتائج تدلل قدرة المصنف العالية للتعلم كلما زاد حجم كمية البيانات.

تمت الاستفادة من هذه الدراسات السابقة في دراستنا الحالية في الآتي

1. الدراسة الاولى: وجود تحليل بيانات والاستفادة منها لاتخاذ قرار مستقبلي كما في حالة القروض هنا هل ثصدق ام لا لجهة ماتكون في حالة الاستيراد لسعات المولدات هل تتم ام لا لزيادة العمليه التسويق
2. الدراسة الثانية: بناء نموزج شجرة القرار بإستخدام خوارزمية C4.5 لتحديد سعات المولدات المستقبلية ذات الشراء العالي وغيرها التي لا يتم شراءها وذلك للاعلان عنها